LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-418161

Dynamics of marine bacterial and phytoplankton populations using multiplex liquid bead array technology

X. Mayali, B. Palenik, R. S. Burton

October 19, 2009

Environmental Microbiology

Disclaimer

Dynamics of marine bacterial and phytoplankton populations using multiplex liquid bead array technology

Xavier Mayali[1], Brian Palenik, Ronald S Burton

Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla CA 92093-0202

[1]Current address: Xavier Mayali, Lawrence Livermore National Laboratory, 7000 East Ave, Livermore CA 94550, tel (925) 423-3892, email: mayali1@llnl.gov

Running title: bacteria and phytoplankton bead array

Summary

Heterotrophic bacteria and phytoplankton dominate the biomass and play major roles in the biogeochemical cycles of the surface ocean. Here, we designed and tested a fast, high-throughput, and multiplexed hybridization-based assay to detect populations of marine heterotrophic bacteria and phytoplankton based on their small subunit ribosomal RNA sequences. The assay is based on established liquid bead array technology, an approach that is gaining acceptance in biomedical research but remains underutilized in ecology. End-labeled PCR products are hybridized to taxon-specific oligonucleotide probes attached to fluorescently coded beads followed by flow cytometric detection. We used ribosomal RNA environmental clone libraries (a total of 450 clones) and cultured isolates to design and test 26 bacterial and 10 eukaryotic probes specific to various ribotypes and genera of heterotrophic bacteria and eukaryotic phytoplankton, respectively. Pure environmental clones or cultures were used as controls and demonstrated specificity of the probes to their target taxa. The quantitative nature of the assay was demonstrated by a significant relationship between the number of target molecules and fluorescence signal. Clone library sequencing and bead array fluorescence from the same sample provided consistent results. We then applied the assay to a 37-day time series of coastal surface seawater samples from the Southern California Bight to examine the temporal dynamics of microbial communities on the scale of days to weeks. As expected, several bacterial phylotypes were positively correlated with total bacterial abundances and chlorophyll *a* concentrations, but others were negatively correlated. Bacterial taxa belonging to the same broad taxonomic groups did not necessarily correlate with one another, confirming recent

results suggesting that inferring ecological role from broad taxonomic identity may not always be accurate.

Introduction

Planktonic microbial communities, with cell numbers on the order of one million cells per ml, play a central role in controlling carbon cycling in the surface ocean (Azam, 1998; Azam and Long, 2001). These communities consist of a diverse assemblage of prokaryotes and eukaryotic protists, often with hundreds or thousands of species present in a single ml of seawater. Understanding the dynamics of these communities requires that we can identify and quantify the abundance of component taxa. Although large phytoplankton cells can often be identified by morphological features, smaller eukaryotes as well as bacteria and archaea are identified primarily by their DNA sequences, typically based on small subunit ribosomal RNA genes (Woese et al., 1985). Even for large phytoplankton, cultivation-independent characterization by rRNA sequencing is now part of the standard methodology to describe organisms (Metfies et al., 2006). The use of these (and other) genes has further led to the design of methods for the rapid characterization of microbial community structure. Some of the most widely used include denaturing gradient gel electrophoresis (DGGE, Muyzer et al., 1993), terminal restriction fragment length polymorphism (TRFLP, Liu et al., 1997), and automated ribosomal intergenic spacer analysis (ARISA, Brown et al., 2005). These methods are able to separate (by electrophoresis) different ribosomal RNA types (ribotypes) based on sequence length or base pair composition and allow rapid fingerprinting of microbial communities for comparison across space and time.

New, and potentially faster and more high-throughput microbial community fingerprinting methods are now being developed, based on competitive hybridization between environmental DNA (or RNA) and target oligonucleotides. One approach utilizes solid microarray technology

that detects successful hybridization with fluorescence (Brodie et al., 2006) or electronic signal (Barlaan et al., 2007). Another approach utilizes liquid bead array technology followed by fluorescence detection by flow cytometry (Spiro et al., 2000). In the latter method, fluorescently-labeled PCR product or nucleic acid extract is hybridized to polystyrene beads that themselves contain different ratios of two fluorescent dyes. Each type of bead is conjugated to a distinct oligonucleotide that acts as a probe for a specific taxon. The mixture is then passed through a flow cytometer able to quantify the amount of hybridized PCR product (or labeled DNA or RNA) and the type of bead. This offers several advantages over solid phase arrays, including favorable liquid hybridization kinetics, the capacity to analyze hundreds of samples in a short amount of time, and the ability to quickly alter the assay (by adding or removing bead types).

Thus far, multiplex liquid array technology has been used in environmental microbiology with PCR for the detection of fungal (Diaz et al., 2006) and bacterial pathogens (Baums et al., 2007). It has also been used directly with extracted RNA to examine metal contaminated soil (Chandler et al., 2006). In marine ecology, this technology has been used for the detection of abundant phytoplankton groups, with direct labeling of extracted DNA and no PCR step (Ellison and Burton, 2005). Here, we report on the development of a PCR-based liquid array method to detect bacteria and eukaryotes in coastal marine samples. We first sequenced 16S and 18S clone libraries from water samples and identified bacterial and phytoplankton target taxa. We then designed and tested probes for those groups in multiplex format, after which we applied the assay on DNA extracts from a time series to illustrate its usefulness for high throughput population dynamics studies.

Results

Clone libraries and probe design. Using universal 16S/18S rRNA primers, a total of 449 clones were sequenced from 4 libraries, comprising 394 bacterial, 1 archaeal, 10 chloroplast, and 44 eukaryotic sequences. This indicates that the universal primers were successful in amplifying all domains of life; the dominance of bacterial sequences over the other domains and their relative abundances seem consistent with the coastal marine surface-water origin of the samples. The bacterial sequence data were dominated by α and γ Proteobacteria as well as cyanobacteria and Bacteroidetes. There were also several sequences from the Verrucomicrobia, Firmicutes, and Actinobacteria groups. Many eukaryotic sequences were similar (or identical) to copepods and dinoflagellates, while some sequences were most similar to uncharacterized eukaryotes from the alveolate and stramenopile groups. Chloroplasts were from diatoms, dinoflagellates, and chlorophytes. Although many taxa, particularly bacteria, were shared among several of the libraries, there were notable differences among the four libraries. While ecologically important, a detailed analysis of these differences is beyond the scope of this report.

Bacterial probes were designed for groups that included at least one of our clones and one or more sequences from Genbank and/or from the Global Ocean Survey metagenomic database. In this study we focused on heterotrophic bacteria as bacterial autotrophs are part of a separate ongoing study (V. Tai R. Burton, B. Palenik, unpublished data). The heterotrophic bacterial taxa targeted in this study can be divided into two general groups. The first included 16S phylotypes identified as being abundant in surface temperate marine waters by previous studies (Brown et al., 2005; Rusch et al., 2007 and others), including members of the SAR11, SAR86, SAR116, *Roseobacter*, *Bacteroidetes*, and *Acidomicrobia* groups. We targeted these ubiquitous and abundant taxa because their numerical dominance suggests they are the primary mediators of

biogeochemical reactions in these environments. The second group of bacterial targets included phylotypes less commonly encountered in rRNA databases but found abundant in our clone libraries from algal bloom waters and from other studies of algal blooms in temperate waters. We targeted this second group because of our interest in algal-bacterial interactions in mediating carbon flux and algal bloom dynamics. Target taxa (from either group) did not always comprise the same level of 16S nucleotide diversity because the degree of 16S diversity is not constant among different phylogenetic groups. For example, one probe might target a group of sequences that share 99% similarity at the 16S level, while another probe might target a group sharing 97% similarity. We report only probes that exhibited a signal to noise ratio over 20 and little to no non-specific signal from clones outside the target group (Table 1). Signal to noise was defined as the ratio of the fluorescence signal from the target clone divided by the fluorescence signal from a negative PCR control reaction. In terms of signal strength, the bacterial probes could be divided into two types. The first, consisting of 10 probes, exhibited acceptable (~20) or better signal/noise and no non-specific signal from tested clones outside the target taxon (an example is shown in Fig. 1a). The second group, consisting of 16 probes, exhibited non-specific signal from two clones or less with a signal of at most 50% of the positive signal (an example is shown in Fig. 1b). An additional 12 probes that exhibited excessive non-specific signal (or no signal with their intended target) are not reported and were discarded from any further analyses. In addition, we designed and tested several probes that showed positive signal with their intended targets but did not show signal from field samples (Table S1).

Due to the short length of the PCR product (desirable for probe hybridization), it was difficult to design more than one probe for each taxon. We successfully achieved this for one of the bacterial target taxa to demonstrate reproducibility and specificity of the assay. Probe #45

specific to clone 2A_D08 (γ Proteobacteria, SAR86 group) demonstrated good signal against its target clone as well as clone 1_E04, which has identical probe binding sequence and is 99% similar over the entire 16S sequence (Fig. 1c). Probe #73 is also specific to clone 2A_D08, located 13 bp downstream, but has a 3 bp mismatch with clone 1_E04. As expected, this probe gives a positive signal with clone 2A_D08 but very little signal with clone 1_E04 (Fig. 1d).

Eukaryotic probes were designed to differentiate phytoplankton genera commonly found off the Southern California coast as well as smaller eukaryotes with cultures available (Table 2). Due to the smaller number of cultures and clones to test specificity, probes that displayed any non-specific signal were not studied further. For example, a probe theoretically specific for the dinoflagellate genus *Alexandrium* was discarded due to nonspecific signal with several other dinoflagellates (data not shown). Successful probes targeted large dinoflagellates (genera *Lingulodinium*, *Scrippsiella*, *Akashiwo*, *Prorocentrum* and *Ceratium*), large diatoms (*Chaetoceros*, *Cylindrotheca*, and *Skeletonema*) and the smaller autotrophic protists *Micromonas and Ostreococcus.* In general, probes were specific at the genus level, although several exceptions occurred, particularly among the diatoms (Table 2).

Sensitivity and specificity. The first step to determine the ability of the method to quantify different targets simultaneously was to mix known quantities of PCR products from single clones before analysis with the Luminex. Clones 2A_F06 and 2D_C12 were amplified separately, their PCR products quantified, and analyzed with the Luminex on their own (including 2D_C12 in two different concentrations), as well as mixed together in equal concentrations. Variability among replicate PCR reactions was low (CV = 7-10%), demonstrating good reproducibility. Luminex signal was consistent whether the target clones were analyzed separately or mixed

together (Fig. 2), and the quantified target of lower concentration exhibited lower Luminex signal, as expected.

The next step to validate the assay for use with mixed community DNA was to investigate the potential to follow the population dynamics of individual target taxa within a mixed assemblage. The experimental design was to make serial dilutions of two clones over a range of concentrations that we would expect to encounter in natural samples. We spiked these dilutions into DNA extracted from a field sample (rather than simply into sterile water) before PCR to mimic conditions that might affect the amplification. This also allowed us to control for well-to-well variation in overall fluorescence that we believe to be caused by variations during the PCR as well as during the hybridization and washing steps of the Luminex assay. For example, variable staining intensity was partially caused by some liquid being left in the wells after washing steps due to the gentle manual pipetting necessary to avoid removing beads. To account for these variations, these standard curves (as well as all field data) were normalized according to the overall fluorescence signal of the well, calculated by adding the Luminex fluorescence values of all the bead colors in each well (see methods).

We also found that amplifying with too many cycles of PCR (>30 cycles, annealing temperature = 52°C) resulted in poor dynamic range of standard curves and potentially overestimated the abundance of rare members of the community (data not shown). Thus, we used 25 cycles of PCR for these standard curves and the field sample analyses. Over the range of 10^4 to 10^8 rDNA copies, the Luminex assay resulted in remarkably consistent reproducibility (CV ranging from 2-12%, with one exception; see below) and a linear relationship between log-transformed target abundance and normalized fluorescence signal (Fig. 3). For one of the clones (2A_F12, Fig. 3b), signal inhibition occurred at the highest concentration tested (10^8 rDNA

copies; CV = 38%). Repeats of this experiment resulted in the same finding, suggesting the presence of a PCR inhibitor in the clone 2A_F12 DNA sample.

After testing specificity and sensitivity of the probes with single clones or mixtures of two clones, the subsequent step was to validate the multiplex Luminex assay with known field samples. The four Scripps Pier water samples originally used to construct clone libraries were analyzed with Luminex but results were not always consistent with sequencing data, with sometime high signal for target taxa that were rare or below detection by clone library sequencing, and vice-versa (data not shown). One caveat of this analysis is that we had relatively few sequences per library (< 150), implying that further sequencing would likely produce more target taxa. Another likely reason, however, was that the PCR primers used for sequencing were not the same as those used for the Luminex assay, as1 kb amplicon was useful for taxonomic identifications but proved to be too large for efficient hybridization in the bead assay. We hypothesized that PCR primer bias could be at least partially responsible for the observed inconsistency between Luminex signal and sequence data. Therefore, we sequenced one additional library (95 clones, from sample 5/21/07) using the same primers as for the Luminex assay (80 bp instead of 1 kb amplicons). We combined the two libraries from sample 5/21/07 and determined which bacterial taxa targeted by the Luminex assay were detected by sequencing and which were not. We considered a targeted taxon present in the sample if any sequences matched the probe sequence with 1 bp mismatch or less (using a less stringent 2 or 3 bp mismatch criterion did not significantly change the analysis). The hypothesis was that if a taxon is not detectable by sequencing, it should have a low Luminex fluorescence signal. Conversely, if a taxon is detectable by sequencing, it should have a higher Luminex signal. The taxa not detectable by sequencing exhibited significantly lower Luminex fluorescence signal than taxa

detectable by sequencing (Fig. 4). There were no false positives (taxa not detected by sequencing with high Luminex signal), while taxa detected by sequencing exhibited a wide range in Luminex signal.

Field samples. Total bacterial counts during the sampling period ranged from 1 to 5 x 10^6 cells mL^{-1} (Fig. 5a) and extracted chlorophyll *a* from 2-12 μg L^{-1} (Fig. 5b). Not unexpectedly, extracted chlorophyll *a* and bacterial abundances were positively correlated (r = 0.54), consistent with bottom-up control of bacterial growth. The data revealed temporal dynamics over the sampling period (Fig. 5c): some taxa were more abundant towards the end of the sampling period while others more abundant towards the beginning (red = high relative abundance, green = low relative abundance). Taxa were grouped by a hierarchical cluster analysis (Fig. 5d) based on the similarity in their abundance patterns over the sampling period. We further performed pairwise correlation analyses to determine taxa with similar (positive correlation) and opposite (negative correlation) distributions over time, as well as correlations with bacterial abundances and extracted chlorophyll *a*. We report correlations with coefficients ≥ |0.4| as Ideker et al. (2001). Fifty-six pairwise correlations had correlation coefficients greater than 0.4 (Fig. 6, highlighted in green), representing taxa with similar temporal distributions. Sixty-six pairwise correlations had correlation coefficients less than -0.4 (Fig. 6, highlighted in red), representing taxa with opposite temporal distributions. Trends of positive and negative temporal interactions existed among bacteria, among phytoplankton, and between bacteria and phytoplankton. Six bacterial taxa were positively correlated with total bacterial abundances and four bacterial and two algal taxa were negatively correlated with bacteria (Fig. 6, column 1). A similar trend was found with extracted chlorophyll *a* (Fig. 6, column 2): eight bacterial and one algal taxon were positively correlated

with chlorophyll, and eight bacterial and three algal taxa were negatively correlated with chlorophyll.

Discussion

Using technology previously established in biomedical research (Dunbar, 2006) and for the detection of bacteria related to coastal water quality (Baums et al., 2007; Tracz et al., 2007), we have developed a hybridization-based assay allowing the detection of bacteria and phytoplankton in marine coastal waters. The assay currently targets 26 bacterial and 10 eukaryotic ribosomal RNA phylotypes, but can be easily expanded as more probes are designed and tested against new targets.

The method described in this report offers several potential advantages for monitoring microbial community dynamics across many samples. First, it targets both bacterial and eukaryotic taxa. To our knowledge, there exists no other molecular fingerprinting method currently used to detect bacterial and eukaryotic microbes concurrently (the most recent version of the Phylochip is an exception; E. Brodie, pers. comm.). Since these two groups of organisms interact very closely in aquatic ecosystems (Cole, 1982; Azam, 1998), such a method is clearly warranted to test ecological questions about their interactions. Further, as we recover more gene sequence data from the marine (and other) environments, the Luminex bead array can be quickly altered by adding one or several new probes to an existing assay. Additional probes can be designed to detect other organisms or more specific groups within currently targeted taxa. Other segments of the 16S rRNA gene can also be amplified to provide different phylogenetic

resolution. This versatility is especially valuable when working with highly dynamic ecosystems such as the coastal ocean, where new information from deep sequencing efforts (Sogin et al., 2006; Rusch et al., 2007) adds to our sequence database on a monthly basis.

Two further advantages offered by the Luminex are high replication and high throughput capabilities. Hundreds of beads of each type (the equivalent of having 100s of identical spots on a microarray) are assayed in every sample, providing statistical accuracy for each probe. The accuracy and reproducibility are also enhanced by the liquid phase kinetics of hybridization that reduce the effects of steric hindrance from solid phase flat arrays (Dunbar, 2006). In addition, a well plate of 96 samples can be assayed (for up to 100 probes) within several hours, providing a large amount of data in a very short time. This combination of reproducibility, multiplex capability, assay versatility, and high-throughput capacity makes this method a potentially useful complement to environmental genomics (Handelsman, 2004; Delong, 2007). Due to the prohibitively high costs of deep sequencing, metagenomics is typically performed on relatively few samples to get a better understanding of the sequence diversity within one ecosystem. However, to constrain hypotheses about ecosystem temporal dynamics (or spatial heterogeneity), a methodology that can assay hundreds of samples in a short period of time is equally valuable. In marine microbiology, the fingerprinting method ARISA (Brown et al., 2005) has revolutionized our understanding of seasonal dynamics (Fuhrman et al., 2006) and latitudinal biogeography (Fuhrman et al., 2008) of planktonic marine bacteria. The Luminex assay offers the opportunity to detect both bacteria and eukaryotes, and can also be altered for the detection of taxa on different taxonomic levels, similar to probes for fluorescent *in situ* hybridization (Pernthaler et al., 2001). In addition, although not carried out here, genes with known biogeochemical functions, such as *nif*H (Moisander et al., 2006) and proteorhodopsin (Beja et al.,

2001) can also be targeted by PCR and the Luminex assay used to provide functional information.

In order to validate the Luminex assay, each individual probe was tested against a suite of pure clones (or cultures) to confirm signal intensity and specificity. Out of over 40 bacterial probes designed to be specific to various phylotypes, 15 probes exhibited too much signal with non-targets and were discarded. *In silico* analyses did not reveal any patterns responsible for this non-specific signal, such as lower numbers of base pair mismatches or higher theoretical melting temperature. Based on this result, future probes should always be tested against both target and non-target DNA before being used on environmental samples. If bead array technology is adopted by additional laboratories, we anticipate that an ever-growing set of probes would become available and investigators could select those of particular interest for their respective analyses while developing new probes as needed. Although the instrumentation employed here can only use 100 different probes at a time, multiple sets of probes can be utilized for a given sample and the new Luminex instrumentation has the capacity of targeting 500 probes.

In considering the relative merits of the bead array approach, it is important to determine the objectives of the analysis. Like all hybridization methods, bead array analysis only reveals taxa for which probes are included, i.e. the coverage of the community is only as complete as the probe set. If a new species invades the system, its presence (even if common) will go undetected unless alternate methods are used to complement the bead array. On the other hand, if the objective is to study the dynamics of specific taxa, the bead array approach appears to be quite viable. Because individual probes are coupled to different bead colors in separate reactions, signal intensity typically varies among different beads when hybridized to equimolar concentrations of their targets (Diaz and Fell, 2004; Chandler et al., 2006). Quantitative

comparisons *among* taxa based on Luminex fluorescence signal will typically require calibration curves for each taxon/probe combination. In our case however, an ANOVA comparing the Luminex fluorescence signal between taxa detected in the clone libraries and those not detected was statistically significant (Fig. 4), suggesting these comparisons are at least semi-quantitative. A more conservative approach is to compare fluorescence signal for a given taxon over many samples to provide direct quantitative data on the population dynamics of individual taxa. We first validated this approach by analyzing standards consisting of pure target clones diluted into a DNA extract from a seawater sample. We tested the ability of the Luminex assay to detect between 10^4 and 10^8 16S rDNA molecules in the PCR reaction (Fig. 3), which resulted in robust quantification. For one clone, however, there appeared to be signal inhibition on the upper end of that scale (10^8 copies; Fig. 3b). Assuming two 16S genes per genome, 100 mL of seawater extracted, and 1/30 of the extracted DNA used in the PCR reaction, this was equivalent to detecting roughly 10^7 bacteria mL^{-1}. Since the maximum total bacterial abundance detected during our sampling period was 5×10^6 mL^{-1}, it appears that Luminex signal inhibition occurred at abundances greater than that expected in our samples, particularly for a single phylotype.

Another aspect of the methodology described here is the data normalization procedure. As described by Chandler et al. in great depth (2006), bead arrays (and all phylogenetic arrays generally) cannot provide absolute abundance data, i.e. comparing the abundance of taxon A versus taxon B in one sample. This is due to the different signal intensities of individual probe-target combinations caused by differences in the base pair composition of the sequence that controls the melting behavior. Therefore, the more conservative approach is to compare the relative abundance of taxon A over many samples, of taxon B over many samples, etc. In this study, the well-to-well variability of the total fluorescence signal was large in some cases,

including the variability between replicates of the same sample. Some wells displayed low signal for all the beads, while other wells displayed high signal for all the beads. We attributed this variability partially to well-to-well differences in the effectiveness of PCR and/or pipetting inaccuracy but more importantly to variation during the post-hybridization washes and staining steps. Artificially high signal can be caused by the staining reagent (streptavidin-phycoerythrin) not being thoroughly washed from a well, and artificially low signal by the staining reagent not being well-mixed into the well. The variation of signal intensity among replicate microarrays is well documented (Spruill et al., 2002) and various normalization procedures are commonly performed to compare them (Do and Choi, 2006). Here, we performed a normalization procedure to account for well-to-well variability in overall signal intensity. We normalized the fluorescence signal of each bead type to the overall fluorescence of the well, the latter calculated by summing the values of all bead types in that well (see equation in methods). Although this approach does not determine changes in the absolute abundances of taxa in the environment, it allows for a meaningful comparison of the relative changes of taxa over time. In other words, the normalized data represent how a given taxon's abundance changes relative to the other taxa. Based on the results of Figures 2 and 3, we are confident in the ability of the Luminex to quantify the amount of amplified target DNA present in a sample, and in our normalization procedure to quantify the relative abundance of targeted DNA across samples. We assume that the DNA extraction efficiency for a given taxon will not change across samples, and any variability in overall DNA extraction efficiency across samples will be removed by the normalization procedure.

Following validation of the method, we analyzed a 37-day time series of surface seawater samples collected from the Scripps pier in Southern California. We uncovered both positive and negative interactions among the Luminex-targeted taxa based on several types of statistical

methods including cross-correlations and cluster analyses. Several bacterial phylotypes were found to correlate positively with both bacterial abundance and extracted chlorophyll *a*. These types of bacteria would likely be considered copiotrophs, fast growers that prefer high organic matter environments (Koch, 2001). Consistent with this idea, we found these sequences in our clone libraries from algal bloom samples but not from our non-bloom library. Furthermore, 16S sequences that match these bacterial targets (Polari-37, Bacter-50, Flavo-63, Roseo-11, and Roseo-19) have been found in previous studies of microbial community structure during algal blooms, including those of diatoms (Riemann et al., 2000; Morris et al., 2006; Rink et al., 2007), dinoflagellates (Fandino et al., 2001; Rooney-Varga et al., 2005), and other phytoplankton types (Zubkov et al., 2002; Brussaard et al., 2005; Barlaan et al., 2007). Conversely, three bacterial phylotypes were found to have negative correlations with bacterial abundance and chlorophyll, including SAR11-13, SAR116-32, and alpha-7. These bacteria would be considered oligotrophs (Koch, 2001), consistent with measured exponential growth rates of 0.7 d^{-1} from laboratory cultures of SAR11 isolates (Tripp et al., 2008).

One result worthy of note was that taxa from different taxonomic groups correlated together rather than with members of the same group (using a cutoff of 0.4 as in Ideker et al., 2001). For example, only one pair among the 4 *Roseobacter* phylotypes were positively correlated (r = 0.52) with one another, one pair among the 5 *Polaribacter* phylotypes (r = 0.52), and no *Rickettsia* phylotypes (Fig. 6). This suggests that using large taxonomic units to infer ecological role may be inaccurate, at least for certain groups. More examination of the population dynamics of phylotypes closely related to one another will be necessary to understand how well 16S similarity can predict ecology in all environments.

The data presented here, obtained using a novel high-throughput method, exemplifies how little is currently known about the dynamics of marine microbial communities over space and time. Using an analysis of many more bacterial taxa than achieved here (171), Fuhrman et al. (2006) found that bacterial communities were seasonally recurring and predictable based on ocean conditions. Since phytoplankton primary production fuels the surface ocean ecosystem, it is not surprising that including these taxa, as done here, provides valuable data to such analyses. Future work in our laboratory will exploit our newly developed assay to uncover temporal and spatial relationships among both bacterial and eukaryotic microbial taxa, many of the former remaining uncultivated and whose ecosystem roles are unknown. Such data can reveal previously uncharacterized interactions that may be an indication of syntrophy between these organisms. This information may become useful to help design optimal conditions for growth in order to isolate and culture these microbes and subsequently uncover their physiology and biogeochemical activities.

Experimental procedures

Clone libraries

Surface water from the Scripps Institution of Oceanography (SIO) Pier (32.86634°, -117.25481°) was collected during (or a few days after) several algal bloom periods, including a period several days after a *Pseudo-nitzschia* diatom bloom (March 23, 2006; SIO pier chlorophyll program), during a *Synechococcus* bloom (May 11, 2006; P. Palenik, unpublished data), and during a mixed species dinoflagellate bloom (May 21, 2007; SIO pier chlorophyll program). To sample

non-bloom communities from a different time of the year, we collected water from eight dates between August and October 2004 and pooled them (after DNA extraction) for a fourth library. Water samples (200 mL) were filtered through 47 mm 0.22 μm pore size polycarbonate filters (Millipore) and frozen at -80°C until extraction. DNA was extracted using a DNeasy Blood & Tissue kit (Qiagen) according to instructions for bacteria. Universal primers (to amplify both eukaryotic and prokaryotic ribosomal RNA genes) were modified from two previous studies (Hovanec et al., 1998; Rivas et al., 2004) using the ARB software (Ludwig et al., 2004): 530F 5'-GTGCCAGCMGCCGCGG-3' and 1390R 5'-CGGGCGGTGTGTRCAARRSSC-3'. A total of 18 cycles of PCR amplification were run to increase sensitivity and minimize PCR artifacts (Acinas et al., 2005), with an annealing temperature of 57°C. Amplified products were cloned with a TOPO® TA Cloning Kit (Invitrogen) and positive clones were sequenced uni-directionally with the M13 forward primer (Agencourt Bioscience). Sequences were trimmed automatically and manually checked using Sequencher (Gene Codes Corp., Ann Arbor, MI). The sequences have been deposited in Genbank under accession numbers EU733720-EU734168. Additional sequences cloned from sample 5/24/07 using Luminex (shorter) primers have been deposited under accession numbers FJ223033-FJ223127. Sequenced clones were frozen at -80°C, regrown in LB (Luria Bertani) broth, and plasmid DNA was isolated with a QIAprep spin miniprep kit (Qiagen). These plasmid samples, containing the partial 16S or 18S rDNA inserts from the clones, were used to subsequently test the probes after PCR (see below). Protist cultures were also used as controls for the eukaryotic probes, and their DNA was isolated as above.

Phylogenetic analysis and probe design

Sequences were added to a ribosomal RNA database in ARB (Jan04 version), which included additional environmental sequences from marine environments (both from Genbank and the Global Ocean Survey, the latter available at the CAMERA website http://camera.calit2.net/). This database included approximately 59,000 aligned sequences in a global phylogenetic tree (data available from X. M.). Clone library sequences were aligned with the ARB internal aligner, manually checked, and added to the global tree using parsimony. Probes were designed with the ARB "probe design" function. This function takes a phylogenetic approach rather than a phenetic one because it groups sequences according to inferred evolutionary relationships rather than simply by sequence similarity. As such, probes were not always specific to the same degree of 16S or 18S sequence diversity (see results and tables 1 and 2). Sequence diversity for each probe was defined as the amount of 16S or 18S diversity among all the taxa matching the probe sequence within one base pair. For bacteria, the 25-bp probes were designed for the region between *E. coli* numbers 967 and 1046, which is a hyper-variable region commonly used for diversity studies (Sogin et al., 2006). We chose such a small region (<100 bp) because initial experiments showed that short PCR amplicons significantly increased fluorescent signal on the luminex flow cytometer (data not shown). This region was also variable enough to differentiate closely related bacterial phylotypes. For eukaryotes, the probes were designed for the region between *E. coli* numbers 1193 and 1380, as the 967-1046 region was not variable enough to differentiate many phytoplankton species. Probes were manufactured with a C-12 spacer at the 5' end (Bioneer Corporation).

Assay development

Probes were conjugated to different colored Luminex xMAP® carboxylated beads (5.6 μ diameter) according to manufacturer's instructions. Each oligonucleotide type is conjugated to approximately 1 million beads in a single reaction. For the bacterial assay, environmental DNA was amplified by PCR (product size ~ 80 bp) with primers modified from a previous study (Sogin et al., 2006) and the forward primers were biotinylated (Table 3). Eukaryotic primers (product size ~ 150 bp) were designed with ARB (Table 3). Amplification was initially performed for 35 cycles with 94°C denaturation (30 s), 52°C annealing (45 s), and 72°C extension (1 min) steps. Subsequently, PCR was decreased to 25 cycles of amplification (see results). Products were checked on agarose gels, and analyzed on a Luminex 100 flow cytometer according to published protocols (Lowe et al., 2004), with modifications. Briefly, amplicons were denatured (95°C) for 5 min and incubated in 1X TMAC buffer (3M tetramethylammonium chloride, 0.1% SDS, 50 mM Tris-HCl [pH 8.0], and 4 mM EDTA [pH 8.0]) at 52°C for 2 hrs with approximately 1000 beads of each color (each color bead carrying a different probe). Incubations were performed in skirted PCR plates covered with plastic film in a thermal cycler. After incubation, samples were washed with fresh TMAC buffer and spun down at 2,000 X g for 3 min. After removing the supernatant, the beads were incubated for 10 min with streptavidin-phycoerythrin (Invitrogen; 250X dilution, in 1X TMAC buffer) in the dark at 52°C, washed, and resuspended in 50 μl 1X TMAC buffer. Data acquisition on the Luminex instrument was performed with Luminex software v 1.7, and a minimum of 50 beads of each color were analyzed. Unless otherwise noted, all Luminex signal values are reported as median fluorescence minus control, the latter defined as the median fluorescence from a negative PCR reaction. In addition, data from field samples and standards spiked in seawater (see below) were normalized to the total array fluorescence of each well This was achieved by summing the fluorescence

values for all the bead colors in each well, and normalizing the value of each bead color to that value with the following calculation:

This procedure was necessary to account for well-to-well variations in overall signal intensity, analogous to variations among replicate microarray analyses. After this normalization procedure, data represent relative rather than absolute abundances in a sample.

Sensitivity and accuracy

The first set of bacterial probes was tested against relatively few clones (~20) to determine optimal hybridization temperature yielding highest signal/noise. Hybridization temperatures between 65 and 45°C were tested every two degrees. After an optimal temperature of 52°C was found, these and all subsequent bacterial probes were tested against 70 different clones from the libraries, including 4 clones of eukaryotic origin. Eukaryotic probes were tested against 17 taxa (a mixture of clones and cultured isolates).

Two experiments were performed to ascertain the quantification potential of the method. The first experiment was to assay mixtures of rDNA from two different clones quantified post-PCR. The number of rDNA copies was calculated based on DNA concentration from purified PCR products (measured with a Nanodrop spectrophotometer, Thermo) and the length of the plasmid plus insert. The second experiment was to determine if the method can detect changes in the abundance of a known target (quantified before PCR) within a mixed sample. This type of spiking experiment more accurately mimics the types of natural samples that we ultimately wanted to be able to analyze. Two different clones were serially diluted over 5 orders of

magnitude into a field sample, PCR amplified, and analyzed as described above (using the normalization procedure). All experimental treatments were performed in triplicate, defined here as separate PCR reactions.

Field sample collection and analysis

In addition to testing the Luminex assay on the 4 samples for which we obtained clone library sequences, we applied the assay to a 37-day time series of surface seawater samples collected from the Scripps pier between March 18 and April 23, 2008. Surface samples were collected between 12:00 and 16:00 daily, filtered onto 47 mm polycarbonate filters (0.22 μm pore size) and frozen at -80°C within 30 min. of collection. DNA was extracted from thawed half-filters as above with a DNeasy Blood & Tissue kit (Qiagen). Duplicate PCR reactions were set up and analyzed with the Luminex assay. Since similar volumes of water were filtered (and then extracted) during the time series, we loaded equal volumes (not concentrations) of DNA extracts in the PCR reactions in an attempt to be as quantitative as possible. Data were background subtracted and normalized to array fluorescence (as described above), and the two replicates were averaged. Cross-correlation and cluster analyses among samples and taxa were performed with the statistical software package JMP v.5.0. Correlation coefficients greater than 0.4 were considered strong as in Ideker et al (2001). Extracted chlorophyll *a* data were obtained from the SCCOOS (Southern California Coastal Ocean Observing System) website (www.sccoos.org). We also quantified total bacterial abundances using flow cytometry. Briefly, a 1 mL subsample was fixed with 0.2 μm-filtered formalin (2% final concentration) and frozen at -80°C. Samples were thawed on ice, duplicates diluted 10 or 100 fold (depending on the sample) in 1X PBS, and

1 stained with SYBRgreen II nucleic acid dye (Invitrogen) for 15 min in the dark. Samples were
2 enumerated with a FACScalibur (BD Biosciences) based on forward scatter and green
3 fluorescence. Controls included stained 1X PBS and unstained seawater samples.
4

Acknowledgements

We thank M. Latz, and M. Hildebrand for protist cultures and P. Huh, C. Tat, and A. Yamamoto for assistance in the laboratory. We are grateful for insightful discussions with V. Tai, E. Brodie, and R. Mueller. This work was supported by a grant from the National Science Foundation to R.S.B. and B.P.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**: 8966-8969.

Azam, F. (1998) Microbial control of oceanic carbon flux: the plot thickens. *Science* **280**: 694-696.

Azam, F., and Long, R.A. (2001) Oceanography - sea snow microcosms. *Nature* **414**: 497-498.

Barlaan, E.A., Furukawa, S., and Takeuchi, K. (2007) Detection of bacteria associated with harmful algal blooms from coastal and microcosm environments using electronic microarrays. *Envir Microbiol* **9**: 690-702.

Baums, I.B., Goodwin, K.D., Kiesling, T., Wanless, D., Diaz, M.R., and Fell, J.W. (2007) Luminex detection of fecal indicators in river samples, marine recreational water, and beach sand. *Mar Poll Bull* **54**: 521-536.

Beja, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411**: 786-789.

Brodie, E.L., DeSantis, T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L. et al. (2006) Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* **72**: 6288-6298.

Brown, M.V., Schwalbach, M.S., Hewson, I., and Fuhrman, J.A. (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Envir Microbiol* **7**: 1466-1479.

Brussaard, C.P.D., Marie, X., van Bleijswijk, J.D., and Veldhuis, M.J. (2005) A mesocosm study of *Phaeocystis globosa* (Prymnesiophyceae) population dynamics. II. Significance for the microbial food web. *Harmful Algae* **4**: 875-893.

Chandler, D.P., Jarrell, A.E., Roden, E.R., Golova, J., Chernov, B., Schipma, M. et al. (2006) Suspension array analysis of 16S rRNA from Fe^{-} and $SO2^{-4}$ reducing bacteria in uranium contaminated sediments undergoing bioremediation. *Appl Environ Microbiol* **72**: 4672-4687.

Cole, J.J. (1982) Interactions between bacteria and algae in aquatic ecosystems. *Ann Rev Ecol Syst* **13**: 291-314.

Delong, E.F. (2007) Sea change for metagenomics? *Nat Rev Microbiol* **5**: 326-326.

Diaz, M.R., and Fell, J.W. (2004) High-throughput detection of pathogenic yeasts of the genus *Trichosporon*. *J Clin Microbiol* **42**: 3696-3706.

Diaz, M.R., Boekhout, T., Theelen, B., Bovers, M., Cabanes, F.J., and Fell, J.W. (2006) Microcoding and flow cytometry as a high-throughput fungal identification system for Malassezia species. *J Med Microbiol* **55**: 1197-1209.

Do, J.H., and Choi, D.-K. (2006) Normalization of microarray data: single-labeled and dual-labeled arrays *Mol Cells* **22**: 254-261.

Dunbar, S.A. (2006) Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clin Chim Acta* **363**: 71-82.

Ellison, C.K., and Burton, R.S. (2005) Application of bead array technology to community dynamics of marine phytoplankton. *Mar Ecol Prog Ser* **288**: 75-85.

Fandino, L.B., Riemann, L., Steward, G.F., Long, R.A., and Azam, F. (2001) Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing. *Aquat Microb Ecol* **23**: 119-130.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104-13109.

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L., and Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105**: 7774-7778.

Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**: 669-685.

Hovanec, T.A., Taylor, L.T., Blakis, A., and DeLong, E.F. (1998) Nitrospira-like bacteria associated with nitrite oxidation in freshwater aquaria. *Appl Environ Microbiol* **64**: 258-264.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K. et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**: 929-934.

Koch, A.L. (2001) Oligotrophs versus copiotrophs. *BioEssays* **23**: 657-661.

Liu, W., Marsh, T., Cheng, H., and Forney, L. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**: 4516-4522.

Lowe, M., Spiro, A., Summers, A.O., and Wireman, J. (2004) Multiplexed identification and quantification of analyte DNAs in environmental samples using microspheres and flow cytometry. In *Environmental microbiology: methods and protocols*. Spencer, J.F.T., and Ragot de Spencer, A.L. (eds). Totowa, New Jersey: Humana press, pp. 51-74.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucl Acids Res* **32**: 1363-1371.

Metfies, K., Töbe, K., Scholin, C., and Medlin, L.K. (2006) Laboratory and field applications of ribosomal RNA probes to aid the detection and monitoring of harmful algae. In *Ecology of harmful Algae*. Granéli, E., and Turner, J.T. (eds). Berlin: Springer-Verlag, pp. 311-325.

Moisander, P.H., Shiue, L., Steward, G.F., Jenkins, B.D., Bebout, B.M., and Zehr, J.P. (2006) Application of a *nif*H oligonucleotide microarray for profiling diversity of N_2-fixing microorganisms in marine microbial mats. *Envir Microbiol* **8**: 1721-1735.

Morris, R.M., Longnecker, K., and Giovannoni, S.J. (2006) Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Envir Microbiol* **8**: 1361-1370.

Muyzer, G., de Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695-700.

Pernthaler, J., Glöckner, F.O., Schönhuber, W., and Amann, R. (2001) Fluorescence in situ hybridization (FISH) with rRNA-targeted oligonucleotide probes. In *Methods in Microbiology*. Paul, J. (ed). London: Academic Press.

Riemann, L., Steward, G.F., and Azam, F. (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol* **66**: 578-587.

Rink, B., Seeberger, S., Martens, T., Duerselen, C.D., Simon, M., and Brinkhoff, T. (2007) Effects of a phytoplankton bloom in a coastal ecosystem on the composition of bacterial communities. *Aquat Microb Ecol* **48**: 47-60.

Rivas, R., Velázquez, E., Zurdo-Piñeiro, J.L., Mateos, P.F., and Molina, E.M. (2004) Identification of microorganisms by PCR amplification and sequencing of a universal amplified ribosomal region present in both prokaryotes and eukaryotes. *Journal of Microbiological Methods* **56**: 413-426.

Rooney-Varga, J., Giewat, M.W., Savin, M.C., Sood, S., M, L., and Martin, J.L. (2005) Links between phytoplankton and bacterial community dynamics in a coastal marine environment. *Microb Ecol* **49**: 163-175.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S. et al. (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: e77.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R. et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA* **103**: 12115-12120.

Spiro, A., Lowe, M., and Brown, D. (2000) A bead-based method for multiplexed identification and quantitation of DNA sequences using flow cytometry. *Appl Environ Microbiol* **66**: 4258-4265.

Spruill, S., J, L., S, H., and B, W. (2002) Assessing sources of variability in microarray gene expression data. *Biotechniques* **33**: 916-920.

Tracz, D.M., Backhouse, P.G., Olson, A.B., McCrea, J.K., Walsh, J.A., Ng, L.-K., and Gilmour, M.W. (2007) Rapid detection of Vibrio species using liquid microsphere arrays and real-time PCR targeting the ftsZ locus. *J Med Microbiol* **56**: 56-65.

Tripp, H.J., Kitner, J.B., Schwalbach, M.S., Dacey, J.W.H., Wilhelm, L.J., and Giovannoni, S.J. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**: 741-744.

Woese, C.R., Stackebrandt, E., Macke, T.J., and Fox, G.E. (1985) A phylogenetic definition of the major eubacterial taxa. *Syst Appl Microbiol* **6**: 143-151.

Zubkov, M.V., Fuchs, B.M., Archer, S.D., Kiene, R.P., Amann, R., and Burkill, P.H. (2002) Rapid turnover of dissolved DMS and DMSP by defined bacterioplankton communities in the stratified euphotic zone of the North Sea. *Deep-Sea Res (II Top Stud Oceanogr)* **49**: 3017-3038.

Table 1: Summary statistics for probes specific for coastal marine bacterial taxa. Probes were tested against a suite of 70 clones of (N = 66) and eukaryotic (N = 4) origin.

probe #[1]	clone name	GB acc.	probe sequence	general taxon	closest GB acc. (% 16S similarity)	probe 16S diversity[2]	signal/ noise	non-specific signal/noise[3]	non-specific clone[4]
3	2B_E02	EU733966	ATTTCTCCAGTTTTTCCCTATATGT	Actinobacteria	AF001652 (99.9)	99	33		
9	2A_H01	EU733936	TCACTAGATTCCCGAAGGCACTCCC	Gamma	AF235120 (99.3)	99	20		
33	2C_H10	EU733850	CAAAGCATCTCTGCTAAGTTTCTAG	Gamma	AF354611 (99.7)	96	37	4	2A_F12
17	2A_F12	EU733923	TTCGAGCACTAAAGCATCTCTGCTA	Beta (OM43)	AY354843 (100)	95	40	6	2A_F09
45	2A_D08	EU733896	TGGTTCCAAAAGGCACACTCTCATT	Gamma (SAR86)	AF001650 (99.9)	98	40	7	2B_A11
7	1_E09	EU734137	GCCGAACTGAAGGTTACCATTCTGT	Alpha	DQ009262 (99.0)	97	23	9	1_C03
13	1_D02	EU734118	CGAAAACTCTAATCTCTTAGAGTCG	SAR11	DQ009203 (99.3)	95	34	7	2C_F10
44	2B_H01	EU733979	TCTGGAATCCGCGACAAGTATGTCA	Roseobacter	AJ400341 (99.6)	99	49	6	2A_A04
11	2B_E05	EU733969	CCATCTCTGGTAGTAGCACAGGATG	Roseobacter	DQ009294 (99.4)	98	39		
19	2D_A01	EU733995	AGCCAGATCTCTCTGGTGGTCATAG	Roseobacter	EU016466 (99.8)	99	30	6	2A_D08, 2A_F09
5	2B_A11	EU733724	ACCAAAATCAGGATGTCAAGACCTG	Rickettsiales	AF406523 (99.4)	98	40	7	2C_H10
32	1_E07	EU734135	TCTCCGGAAACCAAACTCCCCATGT	SAR116	DQ009271 (98.9)	97	75	13	1_D01, 1_E04
25	2A_E11	EU733911	CTAGYCTGTTTCCAAACTATTCGCT	Bacteroidetes	DQ289523 (100)	99	58		
50	2D_C12	EU734027	GAAGAGAAGGCCTGTTTCCAAGCCG	Bacteroidetes	AY274866 (99.6)	99	135		
15	2A_F09	EU733921	AGAAAAGACCATCTCTGATCTATGC	Polaribacter	DQ009115 (99.8)	95	37		
18	2B_B04	EU733954	GATYCATTTCTGAATCATGCAACTT	Polaribacter	AY080916 (99.9)	94	48		
36	2B_H07	EU733768	GGTCTATCTCTAGACCTGTCCCACT	Polaribacter	AJ400347 (99.6)	96	40	10	2B_A11
37	2D_B04	EU734009	ATCTCTAAAGCTGTCAGACTACATT	Polaribacter	AJ400343 (99.9)	92	53	10	2A_H01
40	2A_D09	EU733897	AAGGTCCATCTCTGGTCCATGCAAC	Polaribacter	AM279180 (99.6)	96	54	25	2A_F09[5]
29	2A_F10	EU733922	TTTTCGGGAGAGTACGACAAGCATGT	Verrucomicrobiacea	AY135670 (99.6)	99	47		
54	2B_F01	EU733970	TTTTCCTCTATATGTCAAGCCTTGG	Actinobacteria	AJ575515 (100)	94	30	10	2B_E02
56	2B_G05	EU733759	GTCTCCAGAAACCGTCCTAGGATGT	Alpha	AB378721 (99.9)	94	10		
58	1_F01	EU734139	AGTTCCGAGTATGTCAAGGGGTGG	Gamma	EF574537 (99.2)	95	30	9	2A_A05
60	1_E02	EU734130	CTGAAGAAATTCATCTCTGAAAATC	Rickettsiales	FJ744822 (98.2)	98	18		
63	2B_F11	EU733757	AATAGCTATCTCTARCTAATGCAAC	Flavobacterium	AM279187 (99.5)	90	40	15	2A_H01
67	2A_C09	EU733885	CGTCCAGCCGAACTGAAAGCTCCAT	Roseobacter	FJ826501 (99.3)	95	20	8	2B_H01

[1] probe number is the Luminex Xmap designation

[2] indicates the amount of 16S diversity within the taxon targeted by each probe (1 bp mismatch allowed on probe sequence)

[3] signal to noise ratio of tested clones that do not belong to the target group

[4] clone(s) that gave non-specific signal above the signal from a negative PCR reaction

[5] this clone has a 16S sequence 96% similar to the target clone

1 Table 2: summary statistics of eukaryotic probes

bead color	taxon	other target taxa (no mismatches)	taxa with 1 bp mismatch	sequence	culture/clone used as + control	signal/noise
90	*Emiliania*	*Gephyrocapsa*	none	AAGGTGATAGACTCGTTGAGTGCAT	CCMP374	45
95	*Cylindrotheca*	*Pseudo-nitzschia*	*Pinnularia, Navicula, Craticula, Babesia,*	GGCCAAGGTAGAACTCGTTGAATGC	CCMP343	33
80	*Lingulodinium*	none	*Schizochlamydella*, *Oocystaceae*, *Neochlorosarcina*, *Amphikrikos*	CTTGTTGATCACGTCAGTGTAGCGC	CCMP1932	38
84	*Chaetoceros*	none	*Chaetoceros*	AACACGCGTGCGGTTCAGAACATCT	LD2658	19
65	*Scrippsiella*	none	*Scrippsiella*	ACCCTGCCGGGCAAGCTCATAAACT	local isolate (M. Latz)	13
88	*Micromonas*	*Mantoniella, Ostreococcus*	*Ostreococcus, Bathycoccus*	ACGACGAAATTTGGAGATTACCCAG	CCMP487	20
97	*Pyramimonas*	*Prasinopapilla*	none	GACCCTTCGGCCTAGGTTAGGAGCT	clone 1_B04	15
93	*Ceratium*	*C. tenue*, *C. longpipes*	*C. fusus*, *C. furca*	CCTTCCCAGGACAGGTTAAAGACTC	clone 2A_A04	47
86	*Akashiwo*	none	none	CCTGCCGGACCAGGCAGAAACTCGT	local isolate(M. Latz)	32
82	*Prorocentrum*	none	*Ornithocercus*	GATTTAAAAAGATTACCCAACCCTA	*Prorocentrum* field sample	12

2

Table 3: primers used for Luminex assay to detect bacteria (967F and 1046R) and eukaryotes (1193F and 1380R). Three different bacterial primers (labeled 1-3 each for forward and reverse) were mixed in equal concentrations to eliminate mismatches and used to amplify all bacterial groups. Numbers refer to E. coli 16S nucleotides, and forward primers (denoted by F) were biotinylated.

Primer name	Sequence
Bac967F-1	CAACGCGMARAACCTTACC
Bac967F-2	ATACSCGHRGAACCTTACC
Bac967F-3	ATACGCGAGAAACCTTACC
Bac1046R-1	CGACTYCCATGCTSCACCT
Bac1046R-2	CGACRGCCATGCASCACCT
Bac1046R-3	CGACAGCCATGCAACACCT
Euk1193F	AACAGGTCTGTGATGCCC
Euk1380R	GTGTACAAAGGGCAGGGA

Figure legends

Fig. 1: representative data from multiplex Luminex assay for bacteria using up to 70 clones as targets (arrow indicates target clone); (a) probe # 3 specific for clone 2B_E02 and relatives shows no non-specific signal among the clones tested, (b) probe # 7 specific for clone 1_E09 and relatives shows some non-specific signal with clone 1_CO3 (signal/noise = 9), (c) probe # 45 specific to clone 2A_D08 also hits clone 1_E04 which has identical probe sequence and is 99% similar in 16S sequence, (d) probe # 73 specific to clone 2A_D08 does not give signal with clone 1_E04

Fig. 2: sensitivity analysis of the assay: clones 2A_F06 and 2D_C12 were mixed in various ratios post PCR (number of molecules are indicated below x-axis) and the assay shows consistent signal (median fluorescence minus control, average of three replicates +/- standard deviation).

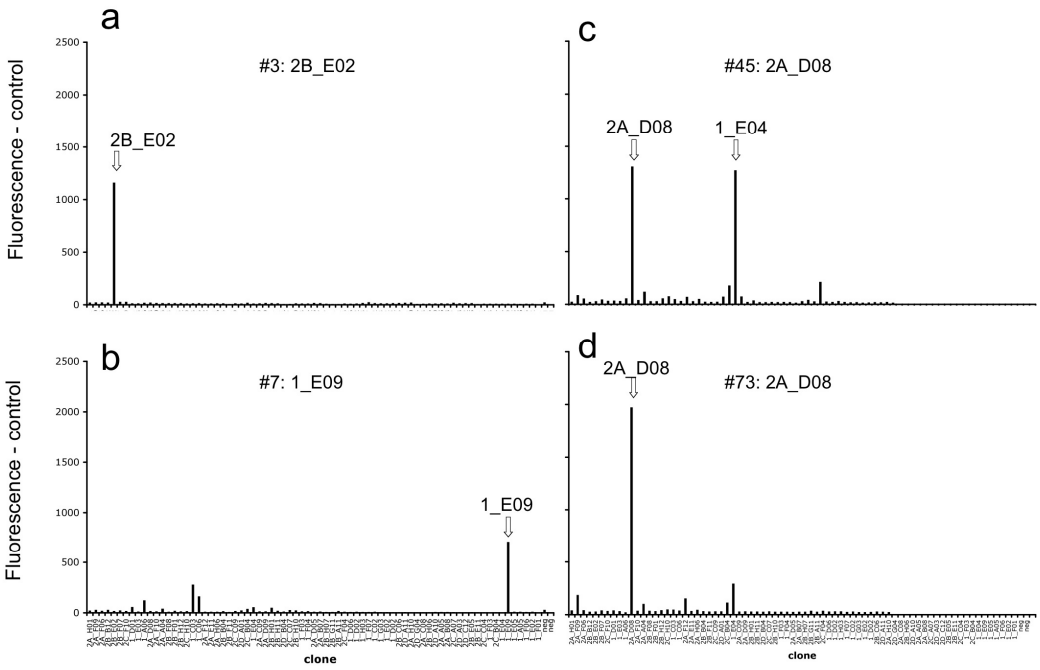
Fig. 3: median Luminex fluorescence signal minus control, normalized to array fluorescence (see text) plotted against the number of target gene copies present before PCR for clones 2D_B04 (a) and 2A_F12 (b) spiked into seawater, showing a significant regression; data represent the average and standard deviations of three replicate PCR reactions.

Fig. 4: Comparison of Luminex fluorescence values (minus control) from sample 5/21/07 between target taxa found by clone library sequencing using two different primer sets versus those that were not found, showing mean and 95% C.I. (gray diamonds). A one-way ANOVA

was significant ($p = 0.0018$), showing that taxa not found by sequencing exhibited lower Luminex fluorescence values.

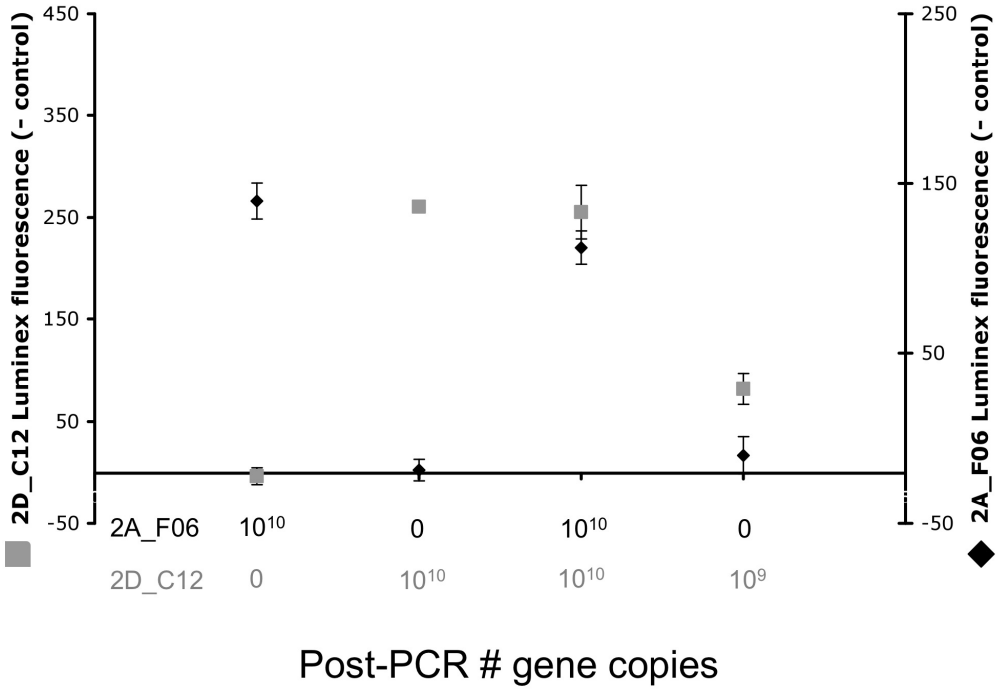
Fig. 5: Color map representing temporal dynamics of 36 Luminex targeted taxa over the course of a 37-day time series (going from left to right) sampled from Scripps pier. Data have been color-coded from low (green) to medium (black) to high (red) abundance for each taxon. Taxa are grouped together (left) based on a hierarchical cluster analysis. Chlorophyll *a* and bacterial abundances from flow cytometry counts are also plotted above.

Fig. 6: Pairwise correlations among Luminex-targeted taxa with correlation coefficients greater than 0.4 (positive correlation, highlighted in green) or less than -0.4 (negative correlation, highlighted in red). Data are based on 37-day time series of surface water collected at Scripps pier between March 18 and April 23, 2008.


a
#3: 2B_E02
2B_E02
Fluorescence - control
c
#45: 2A_D08
2A_D08
1_E04
b
#7: 1_E09
1_E09
Fluorescence - control
d
2A_D08
#73: 2A_D08
clone
clone

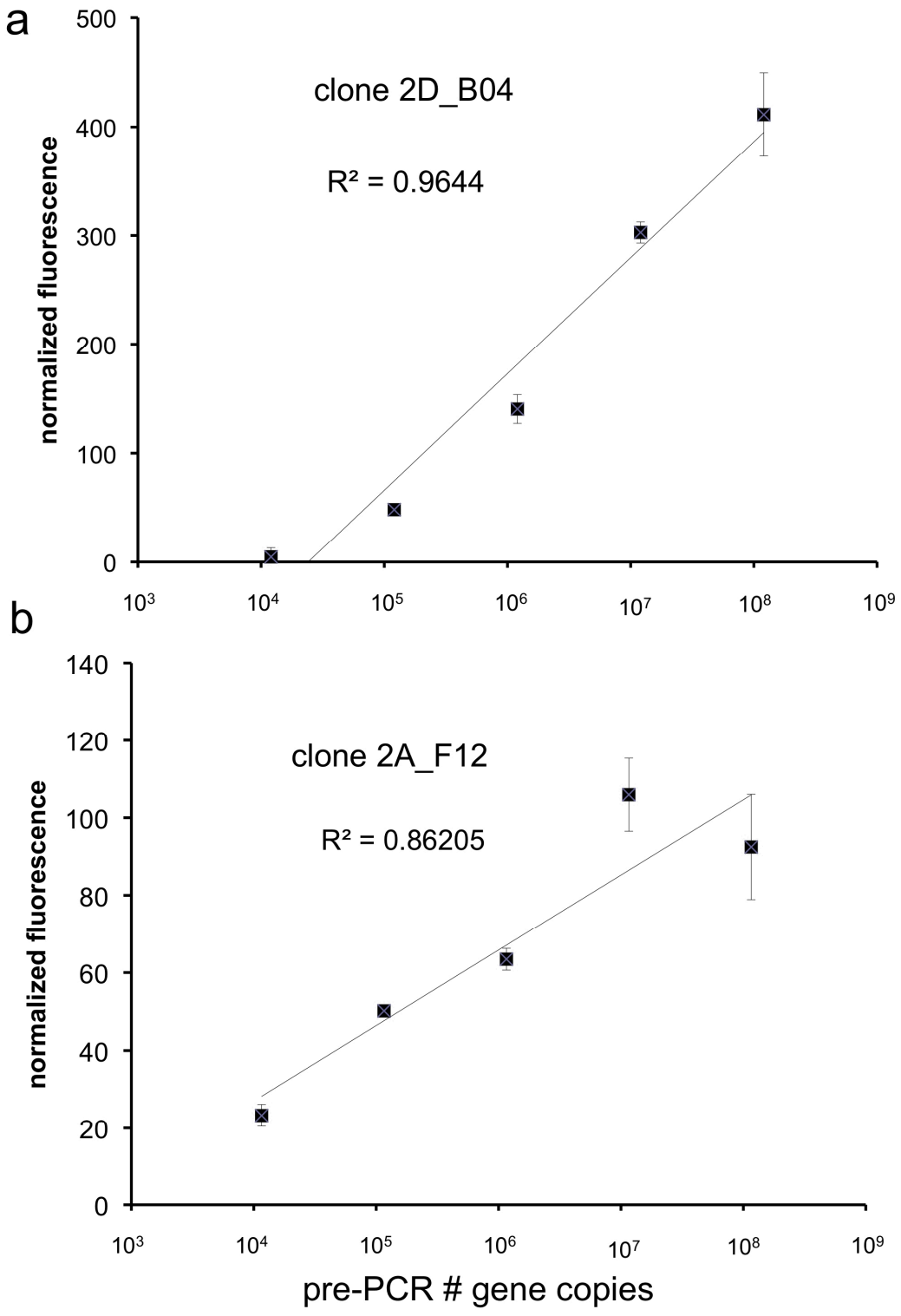
1

2 Figure 1


2D_C12 Luminex fluorescence (- control)
450
350
250
150
50
-50
250
150
50
-50
2A_F06 Luminex fluorescence (- control)
2A_F06 10^10 0 10^10 0
2D_C12 0 10^10 10^10 10^9
Post-PCR # gene copies

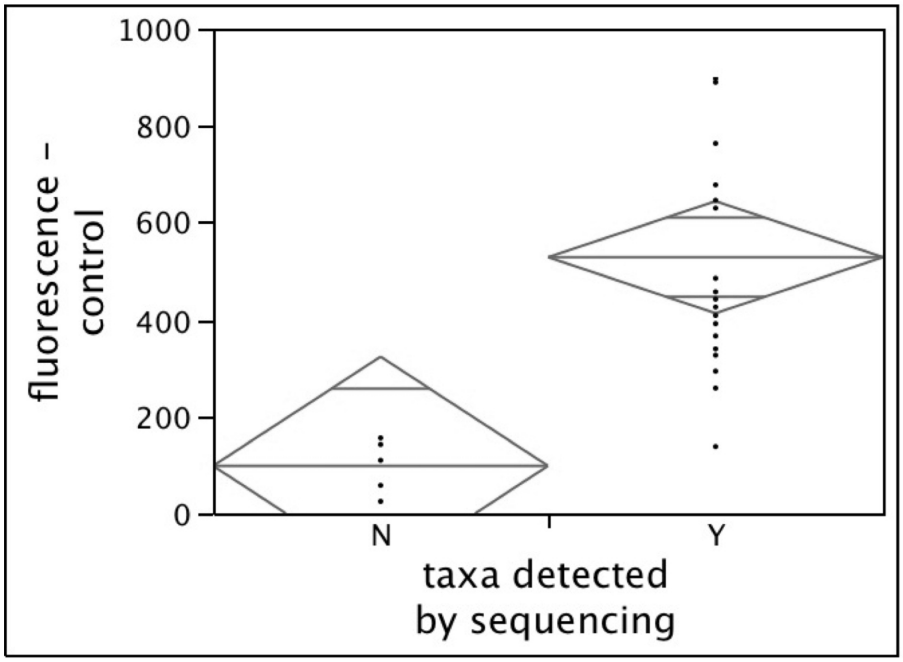
1

2 Figure 2


a
clone 2D_B04
R² = 0.9644
normalized fluorescence
500
400
300
200
100
0
10³
10⁴
10⁵
10⁶
10⁷
10⁸
10⁹
b
clone 2A_F12
R² = 0.86205
normalized fluorescence
140
120
100
80
60
40
20
0
10³
10⁴
10⁵
10⁶
10⁷
10⁸
10⁹
pre-PCR # gene copies

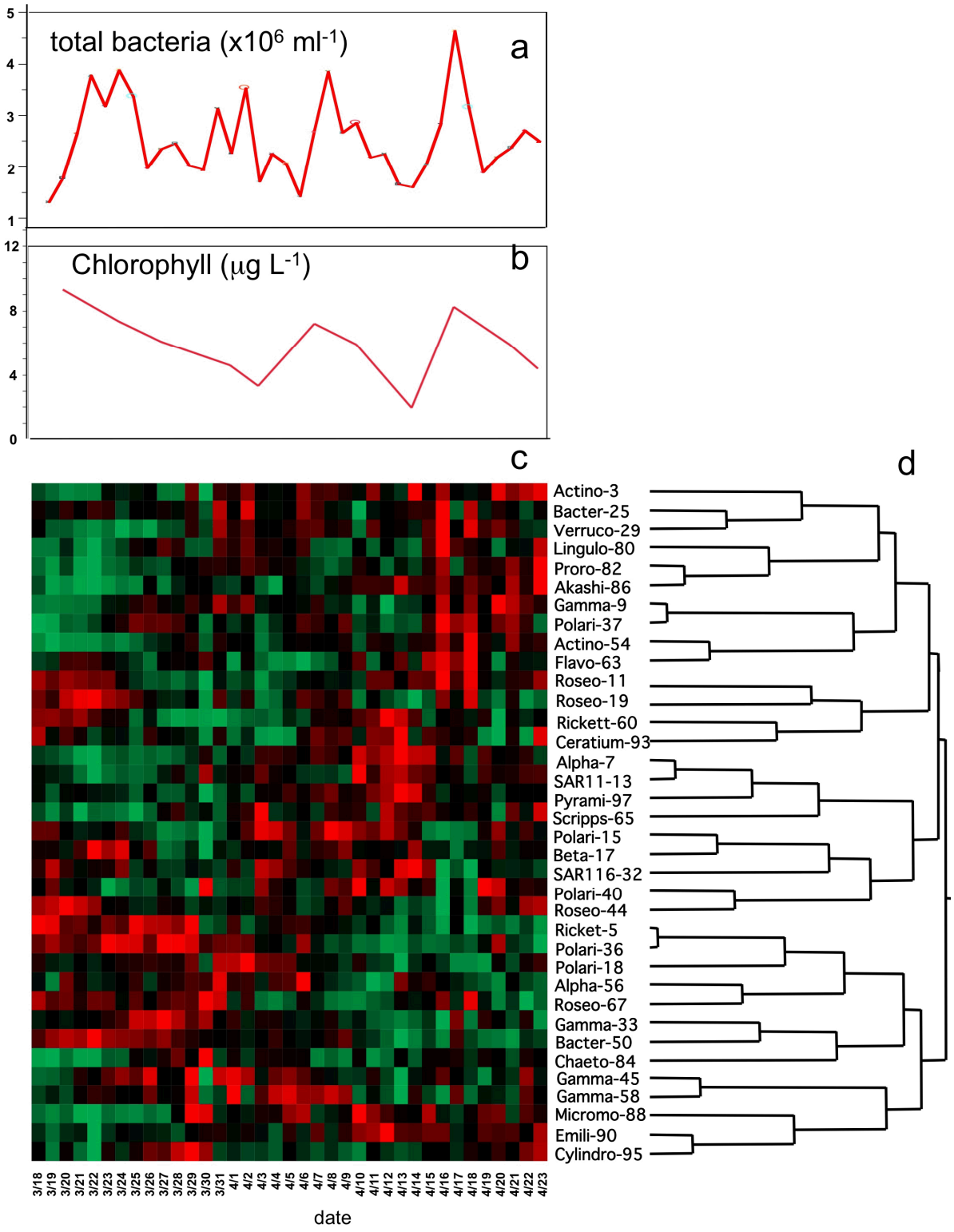
1

2 Figure 3


1000
800
600
400
200
0
fluorescence –
control
N
Y
taxa detected
by sequencing

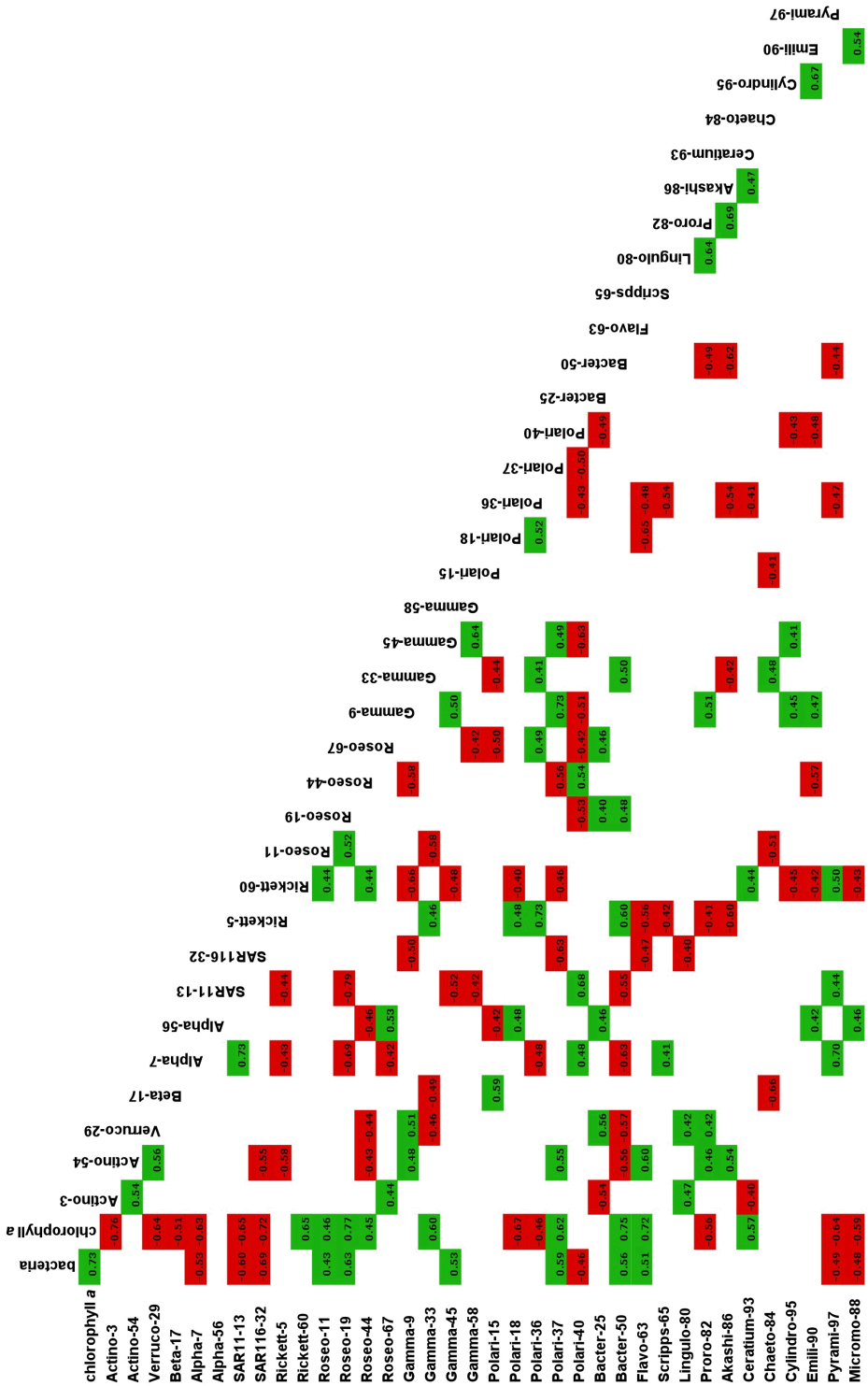
1

2 Figure 4


total bacteria (x10^6 ml^-1)
a
Chlorophyll (μg L^-1)
b
c
d
Actino-3
Bacter-25
Verruco-29
Lingulo-80
Proro-82
Akashi-86
Gamma-9
Polari-37
Actino-54
Flavo-63
Roseo-11
Roseo-19
Rickett-60
Ceratium-93
Alpha-7
SAR11-13
Pyrami-97
Scripps-65
Polari-15
Beta-17
SAR116-32
Polari-40
Roseo-44
Ricket-5
Polari-36
Polari-18
Alpha-56
Roseo-67
Gamma-33
Bacter-50
Chaeto-84
Gamma-45
Gamma-58
Micromo-88
Emili-90
Cylindro-95
3/18 3/19 3/20 3/21 3/22 3/23 3/24 3/25 3/26 3/27 3/28 3/29 3/30 3/31 4/1 4/2 4/3 4/4 4/5 4/6 4/7 4/8 4/9 4/10 4/11 4/12 4/13 4/14 4/15 4/16 4/17 4/18 4/19 4/20 4/21 4/22 4/23
date

1

2 Figure 5


chlorophyll a
Actino-3
Actino-54
Verruco-29
Beta-17
Alpha-7
Alpha-56
SAR11-13
SAR116-32
Rickett-5
Rickett-60
Roseo-11
Roseo-19
Roseo-44
Roseo-67
Gamma-9
Gamma-33
Gamma-45
Gamma-58
Gamma-58
Polari-15
Polari-18
Polari-36
Polari-37
Polari-40
Bacter-25
Bacter-50
Flavo-63
Scripps-65
Lingulo-80
Proro-82
Akashi-86
Ceratium-93
Chaeto-84
Cylindro-95
Emili-90
Pyrami-97
Micromo-88
bacteria
chlorophyll a
Actino-3
Actino-54
Verruco-29
Beta-17
Alpha-7
Alpha-56
SAR11-13
SAR116-32
Rickett-5
Rickett-60
Roseo-11
Roseo-19
Roseo-44
Roseo-67
Gamma-9
Gamma-33
Gamma-45
Gamma-58
Polari-15
Polari-18
Polari-36
Polari-37
Polari-40
Bacter-25
Bacter-50
Flavo-63
Scripps-65
Lingulo-80
Proro-82
Akashi-86
Ceratium-93
Chaeto-84
Cylindro-95
Emili-90
Pyrami-97

1

2 Figure 6

Table S1: List of probes successfully tested against target and non-target clones (or cultures) but omitted from field analyses due to lack of signal in field samples, too few bead colors available for multiplex analysis, or duplicate probe.

Clone (or genus)	Sequence	Taxon	Genbank	Representative GB accession
Pteridomonas	CCCGGCCAAGGTTCTTATACTTGTT	bacterivore	n/a	L37204
Skeletonema	AGTTTGATGAACTGCGATTACTAGG	diatom	n/a	X52006
2A_F06	TTCCAGAAGACATCACTGTGGATTT	Bacteroidetes	EU733918	EU799145
1_C06	TCTCCAAAATCCAAACTACCATGT	Rickettsia	EU734110	DQ009276
2BE11	TGTCACTATGTCCCGAAGGAAAGCC	Roseobacter	EU733750	AY353560
2A_C02	GGCACTGCTTCATTACAAAGCATTC	SAR86	EU733878	AY033328
2A_D08	GCACACTCTCATTACAAGAGCCTCC	SAR86	EU733896	AF001650